



Queensland University of Technology
Brisbane Australia

This is the author's version of a work that was submitted/accepted for publication in the following source:

Zheng, Zuduo, Ahn, Soyoung, & Monsere, Christopher M. (2010) Impact of traffic oscillations on freeway crash occurrences. *Accident Analysis and Prevention*, 42(2), pp. 626-636.

This file was downloaded from: <http://eprints.qut.edu.au/41883/>

© Copyright 2010 Elsevier

Notice: *Changes introduced as a result of publishing processes such as copy-editing and formatting may not be reflected in this document. For a definitive version of this work, please refer to the published source:*

<http://dx.doi.org/10.1016/j.aap.2009.10.009>

1 IMPACT OF TRAFFIC OSCILLATIONS ON FREEWAY CRASH OCCURRENCES

1.1 Introduction

The safety implications of oscillations have been studied largely by evaluating the significance of speed deviations on the likelihood of crash occurrence, though most existing studies did not make an explicit distinction between congested and uncongested traffic. Some early studies (Lave 1985; Solomon 1964) report that larger speed deviations increase the probability of certain types of crashes (e.g., rear-end crashes). Many researchers have challenged this finding; Davis (2002) provides contradictory empirical evidence. Debates continue today, as existing studies exhibit shortcomings in data resolution and methodology. Moreover, only a few studies (Noland and Quddus 2005; Shefer 1997) have investigated the characteristics of crashes in traffic congestion, and they used low-resolution traffic data (e.g., hourly or daily flow), which do not capture oscillatory driving conditions.

The present study seeks to understand the impact of traffic oscillations on the likelihood of freeway crash occurrences by analyzing event-based crash data and high-resolution (20 seconds) traffic data on a freeway segment. Oscillations are measured as variations in congested flow, speed and occupancy (a dimensionless measure of density) over a certain period. A matched case-control design is adopted in which a case corresponds to the traffic condition prior to a crash. The matched control corresponds to the conditions in the same location on a day when a crash did not occur after controlling for geometry, weather and the presence of congestion. The exposures of interest are the

average traffic state (e.g., average speed) and magnitude of oscillations (e.g., speed variations). Conditional logistic regression models were developed based on samples of cases and matched controls. The modeling results show not only that speed variations have a significant impact on (rear-end) crash occurrences but that these variations are more significant than average traffic states.

This chapter is organized as follows. The following section discusses previous research related to the present study. Chapter 2.3 describes in detail the features of the study site and the efforts to process the crash and traffic data. The design of the case-control study and the subsequent modeling efforts are explained in chapters 2.4 and 2.5, respectively. Chapter 2.6 evaluates and interprets the models.

1.2 Background

Three types of relevant literature are reviewed. The author first discusses previous studies that examined the impacts of speed variance and/or congestion on traffic safety. The review revealed that contradictory results have been reported primarily due to limitations in data resolution and analysis methodology. In light of this finding, various methods to relate crash characteristics to traffic features are reviewed. Finally, the author provides a brief background of the case-control design and a rationale for selecting this method for the present study.

1.2.1 Impact of Speed Variance and Congestion on Traffic Safety

Solomon (1964) appears to be the first study to report that speed variations have a larger impact on crash occurrences than average speed. However, his findings received little attention until Lave (1985) claimed in a controversial paper that once one controls

for speed variations, average speed has little impact on highway safety. Several studies (Davis 2002; Fowles and Loeb 1989; Lave 1989; Levy and Asch 1989; Synder 1989) followed, debating Lave's claim. Levy and Asch (1989) found that the mean and variance of speeds were actually correlated, thus disputing Lave's finding. Synder (1989) distinguished between fast and slow vehicles, examining the significance of the speed variance in each group and comparing the average speeds between the two groups. He found that average traffic speed is an important determinant of highway fatalities and that speed variance is important for fast vehicles only. More recently, Malyshkina and Mannering (2008) conducted a before-and-after study to assess the severity of accidents in relation to the increased speed limits (65 to 70 mph (105 to 113 km/h)) in 2005 on some rural interstates in Indiana. They found that the increased speed limits did not have a statistically significant effect on the severity of accidents.

The aforementioned studies used data that were highly aggregated temporally and spatially (e.g., annual statewide data) and may be subject to an ecological fallacy, as noted by Davis (2002). Moreover, no distinction was made between congested and uncongested traffic, even though traffic properties in the two regimes are not the same and may have different impacts on traffic safety.

Fewer studies have examined the effect of traffic congestion on crash occurrences. Shefer et al. (1997) suggest that crash frequency increases in congestion due to increased interactions among vehicles, while crash severity decreases because of the lower speeds of congested traffic. However, they used simulated data to test the performance of their model, which undermines its validity. Noland and Quddus (2005) report conflicting

results based on the analysis of enumeration district data for London, UK, which imply that congestion in urban areas is unlikely to reduce crash severity and frequency. Wang et al. (2009) confirm this finding using the congestion index (CI), which is defined as the ratio of difference in actual and free-flow travel times to the free-flow travel time, to measure the congestion level of a motorway near London, UK. However, the congestion index was calculated from hourly data averaged over a year. Thus, the data resolutions of these studies were not adequate to analyze the effect of oscillatory traffic flow in congestion.

It is evident that existing studies report inconsistent findings on the effect of speed variance and congestion on crashes; this seems attributable to limited data resolution and analytical methods. Moreover, it appears that there has been limited effort, if any, to investigate the effect of traffic oscillations on freeway traffic safety.

In the following section, the author briefly reviews methods to relate crash characteristics to traffic features in order to adopt a statistically sound methodology.

1.2.2 Methods to Relate Crash Characteristics to Traffic Features

For decades, researchers have used various methods to relate crash characteristics to traffic features. Many pioneering studies tried to establish a connection between crash and traffic conditions based on aggregated data (e.g., Lave 1985; Solomon 1964). In essence, they related aggregated indicators of traffic conditions, such as speed deviations (e.g., the annual 85th percentile speed minus the annual average speed), to crash frequency. As mentioned previously, using highly aggregated data, such as annually averaged data, can result in biased findings that are open to interpretation.

Another common method uses the speeds of individual vehicles to compare a crash-involved vehicle to other vehicles (e.g., Cirillo 1968; Research Triangle Institute 1970). This method imposes a great challenge on data collection endeavors, as the trajectories of vehicles in crashes are necessary to measure pre-crash speeds (Xin et al. 2008).

Some researchers have attempted to use individual drivers' accident histories (Fildes and Rumbold 1991; Tilden et al. 1936) by sampling the speeds of selected drivers at a particular location and then analyzing their crash histories. This approach has been criticized for selection bias in the data collection as well as privacy issues (Kloeden et al. 1997).

Finally, some traffic safety studies borrow the case-control design pioneered in the field of epidemiology (Kloeden et al. 1997; Abdel-Aty et al. 2005). This method is suitable for safety-related research, as a crash is a rare event. The next section discusses case-control design in greater detail.

1.2.3 Review of the Case Control Design

The case-control design is an efficient method to study rare events that is particularly prevalent in epidemiology (Manski 1995; Manski 1995; Schlesselman and Stolley 1982) due to its simplicity, cost-effectiveness, and theoretical soundness. The central idea of the case-control study is to compare two groups, one with the outcome of interest (such as disease, death, crash) and one without it, by incorporating potential explanatory factors (exposures) (Cornfield et al. 1959; Cornfield 1951; Fisher 1958a; Fisher 1958b). Cornfield (1951) proved theoretically the validity of the case-control method and demonstrated its effectiveness using epidemiological examples. The theoretical details of

the case-control method are beyond the scope of this paper; interested readers should refer to (Breslow and Day 1980; Cornfield et al. 1959; Cornfield 1951; Schlesselman and Stolley 1982) for more in-depth discussions.

The most natural way to use a case-control design to study traffic safety is at the individual vehicle level. A vehicle involved in a crash is taken as a case, and other vehicles at the crash scene or in similar situations (but not involved in a crash) are taken as controls. The first notable case-control study in the traffic literature is Kloeden et al.(1997), which examined the relationship between speed in uninterrupted traffic flow and the risk of a casualty crash. More recently, Davis et al. (2006) used a case-control design to investigate the effect of speed in run-off-road crashes.

Although employing a case-control design at an individual vehicle level is ideal, collecting data for individual vehicles is difficult in practice. For example, the pre-crash speed of a vehicle is usually estimated using crash reconstruction methods. However, these reconstruction methods are often quite complex, and the estimated speed is subject to bias, as noted by (Davis et al. 2006). In an effort to remedy this shortcoming, Abdel-Aty et al. (2005) developed a case-control design using 5-minute aggregated data from inductive loop detectors to develop a logistic model for multi-vehicle crashes. However, they did not distinguish un-congested and congested traffic regime and did not focus on oscillations in congestion.

Of note, some studies have employed case-control designs to examine the safety impact of certain risk factors. Gross and Jovanis (2007) provide a review of case-control studies used in the highway safety literature. They also developed a matched case-control

design to quantify the effect of road geometry such as lane and shoulder widths.

However, traffic operational features (oscillations in particular) were not examined in their study.

In the present study, a case-control design is implemented in combination with event-based crash data and high-resolution (20-second) traffic data to examine the impact of traffic oscillations in congestion on the likelihood of crash occurrence.

1.3 Study Site and Data Processing

The study site was a 12-mile (19-km) section (mileposts 296.26 to 307.9) in the northbound direction on I-5 in Portland, OR (see Figure 1 for a schematic). This location was selected among other freeway corridors in the region based on several criteria: extent of congestion and oscillations, spacing of loop detector stations, traffic data quality, crash data availability, and crash sample size. Note that this segment contains three lanes, and from mileposts 303 to 306 the left-most lane is dedicated to High Occupancy Vehicles (HOV). The HOV lane is limited to vehicles with at least two passengers from 3 to 6 p.m. on weekdays. Spacings between loop detector stations range from 0.34 to 2.37 miles (0.6 to 3.9 km) with an average of about 1.06 miles (1.7 km). Data from the loop detectors (vehicle count, occupancy and time-mean speed during each 20-second interval) are available from Portland Oregon Regional Transportation Archive Listing (PORTAL 2009).

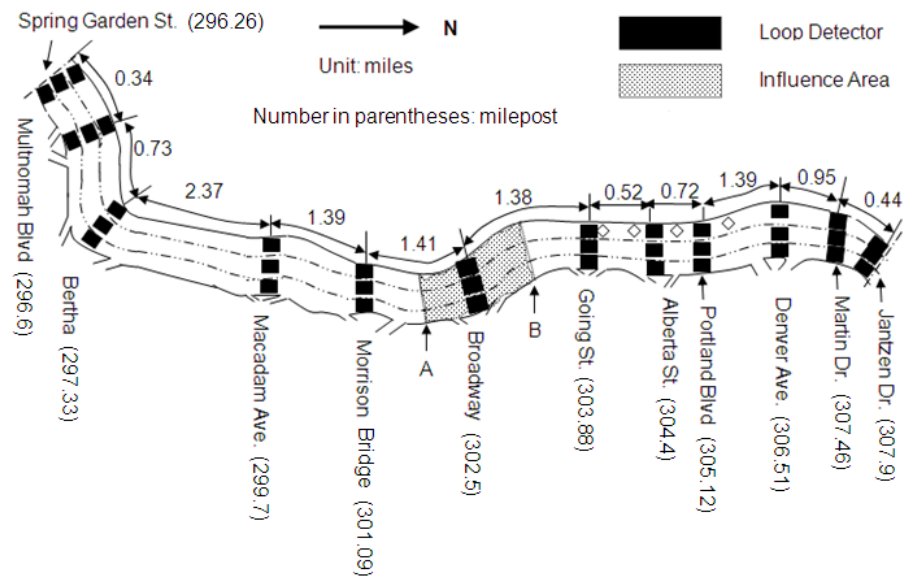


Figure 1.1 Schematic of the study site, northbound I-5, Portland, OR

Loop detectors (black squares in Figure 1.1) measure traffic conditions at 12 locations on the selected freeway segment. Notably, “influence areas” are defined that are bounded by the midpoints between neighboring stations. The shaded area in Figure 1 is an example of an influence area. *A* corresponds to the midpoint between Broadway St. and Morrison Bridge, and *B* corresponds to the midpoint between Broadway St. and Going St. Table 1.1 provides more detailed information on the influence areas of the study segment.

Table 1.1

Mileposts of loop detector stations and influence areas

ID	Milepost	Location Description	Influence Area
1	296.26	Spring Garden St.	(295.72 - 296.43)
2	296.6	Multnomah Blvd.	(296.43 - 296.97)
3	297.33	Bertha Blvd	(296.97 - 298.52)
4	299.7	Macadam Ave.	(298.52 - 300.40)
5	301.09	Morrison Bridge	(300.40 - 301.80)
6	302.5	Broadway St.	(301.80 - 303.19)
7	303.88	Going St.	(303.19 - 304.14)

8	304.4	Alberta St.	(304.14 - 304.76)
9	305.12	Portland Blvd.	(304.76 - 305.82)
10	306.51	Denver Ave.	(305.82 - 306.99)
11	307.46	Martin Dr.	(306.99 - 307.68)
12	307.9	Jantzen Dr.	(307.68 - 307.9)

It is assumed that traffic conditions (including oscillations) are the same within each influence area and are represented by the measurements taken at the corresponding detector station. This assumption was made due to the inherent limitation of loop detectors, which measure traffic conditions at specific points rather than over space. Moreover, the exact locations of crashes are unknown, and thus, it is not straightforward to estimate oscillations at the exact crash location. Nevertheless, it is found that the assumption of uniform oscillations within an influence area is quite reasonable as demonstrated later in this section.

The speed contour plot in Figure 1.2 illustrates the average traffic conditions of the study section during weekdays in January 2006. Time and space (in terms of mileposts) are shown on the x- and y-axes, respectively, and the color scale represents the estimated time-mean speeds over the month according to the legend in the figure. A distinct congestion pattern, which is marked by the dark time-space regions, occurs during p.m. peak hours (defined as 4:00 – 6:00 p.m.). Congestion starts around 3:00 p.m. and ends around 6:00 p.m. The bottleneck for this congestion is located downstream of milepost 307, and the resulting queue propagates to about milepost 296. Moreover, congestion is present recurrently on this segment such that more than 85 percent of the p.m. peak hours on weekdays exhibit congestion on this freeway stretch.

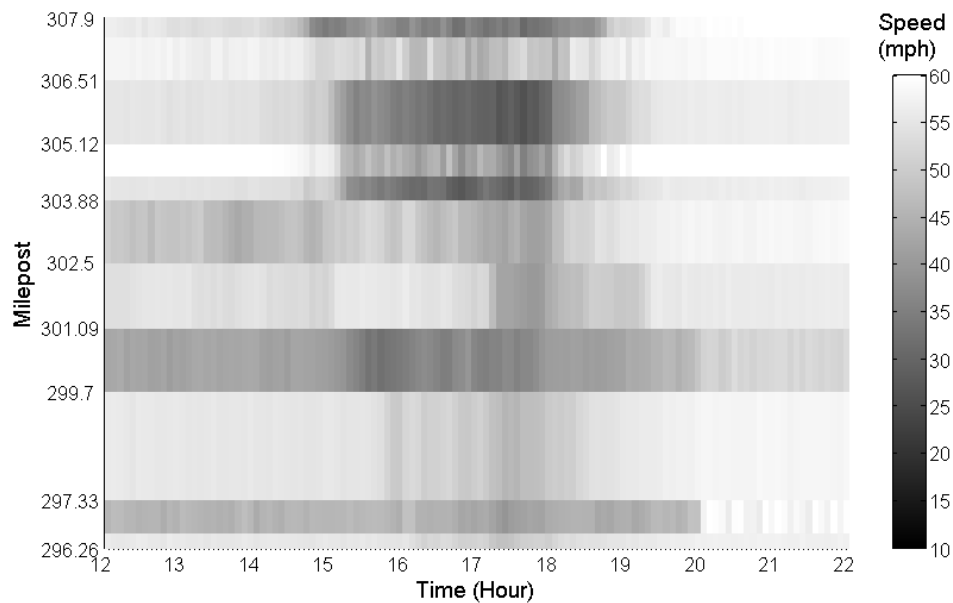


Figure 1.2 Speed contour for the study segment: Weekdays in January 2006

During the congested periods, traffic oscillations were examined visually using time-series speed plots to confirm their presence and amplitudes. For instance, Figure 1.3 shows speed over time (4:30 – 5:30 p.m.) at Jantzen Dr. on May 14, 2007. The plot exhibits a recurring pattern of acceleration followed by deceleration with the speed ranging from around 12 to 37 mph (19 to 60 km/h). The figure also shows varied oscillation amplitudes such that some periods exhibit larger speed variations (e.g., 4:50 – 5:00 p.m.) than others (e.g., 5:20 – 5:30 p.m.).

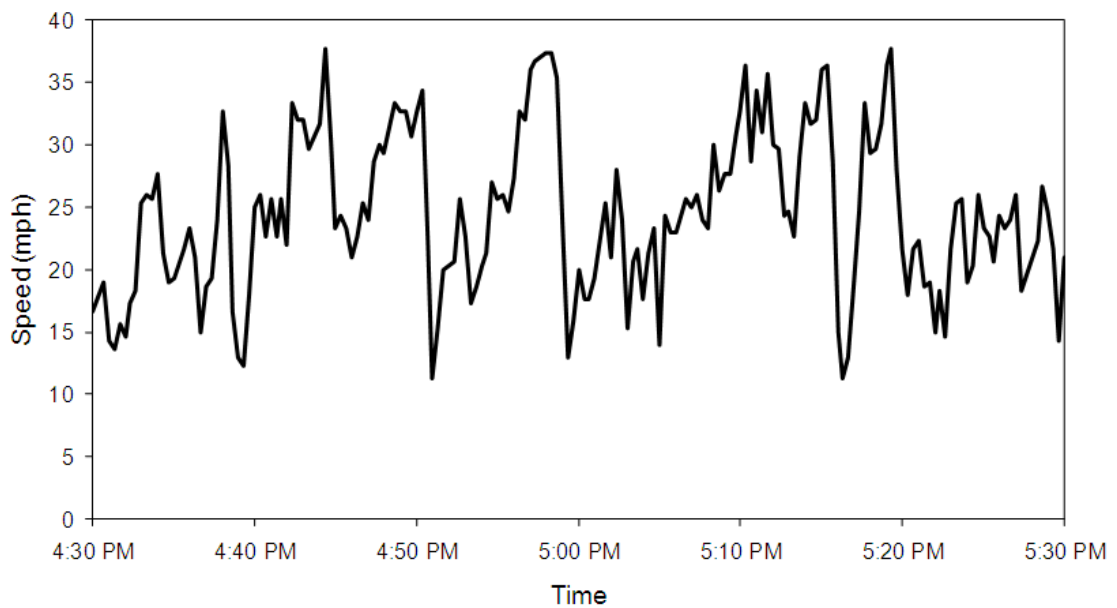


Figure 1.3 Speed time series plot at milepost 307.9 on May 14, 2007

Traffic data from loop detectors often contain invalid data (e.g., missing value, negative value, and non-zero speed with zero count) due to malfunctioning detectors, communication failures, and other reasons. Figure 1.4 presents monthly percentages of valid traffic data during p.m. peak hours in 2004-2007 for each detector station. The monthly percentages of valid data are shaded according to the color scale shown in the legend: darker regions correspond to lower data quality; traffic data for 2004 were available only in September, October, and November. The figure indicates that the percentage of valid data is reasonably good (mostly larger than 90 percent) for the selected study segment. Note that days with invalid traffic data prior to crashes were excluded in the further analysis.

Crash-related data are obtained from two databases maintained by the Oregon Department of Transportation (ODOT); the statewide Crash Data System (CDS) contains all reported crashes on public roads in Oregon and the incident database records all

freeway incidents in the Portland Metropolitan area as logged by the Traffic Management and Operations Center (TMOC). Crashes in the CDS are recorded to the nearest hour while the incident data is at much higher resolution (minutes). CDS was used to identify crashes, and the incident data were used to confirm the crashes and retrieve more detailed information about the crashes. A total of 242 crashes were reported in the study segment during the evening peak periods in 2004 to 2007. Several steps were taken to obtain or estimate the occurrence times of these crashes. The occurrence times were extracted from the incident database since the incident database provides more accurate incident occurrence times than the crash database. For crashes missing from the incident database, 20-second traffic data were used to estimate the occurrence times based on the assumption that a crash in queued traffic will be reflected in the traffic characteristics (e.g., a sharp speed drop (increase) at an upstream (downstream) detector station).

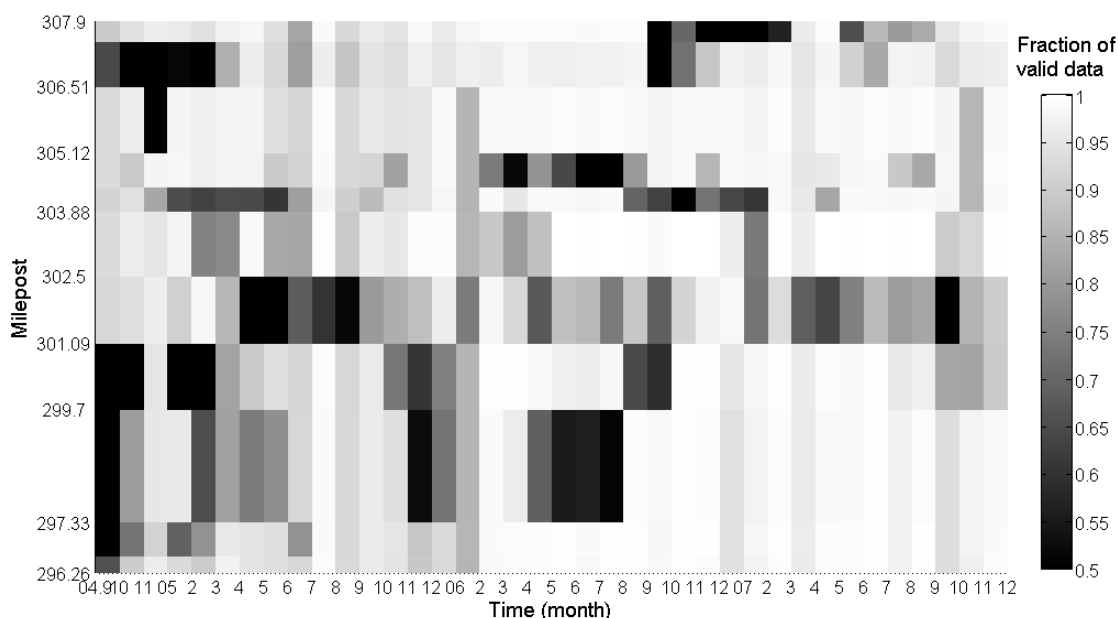
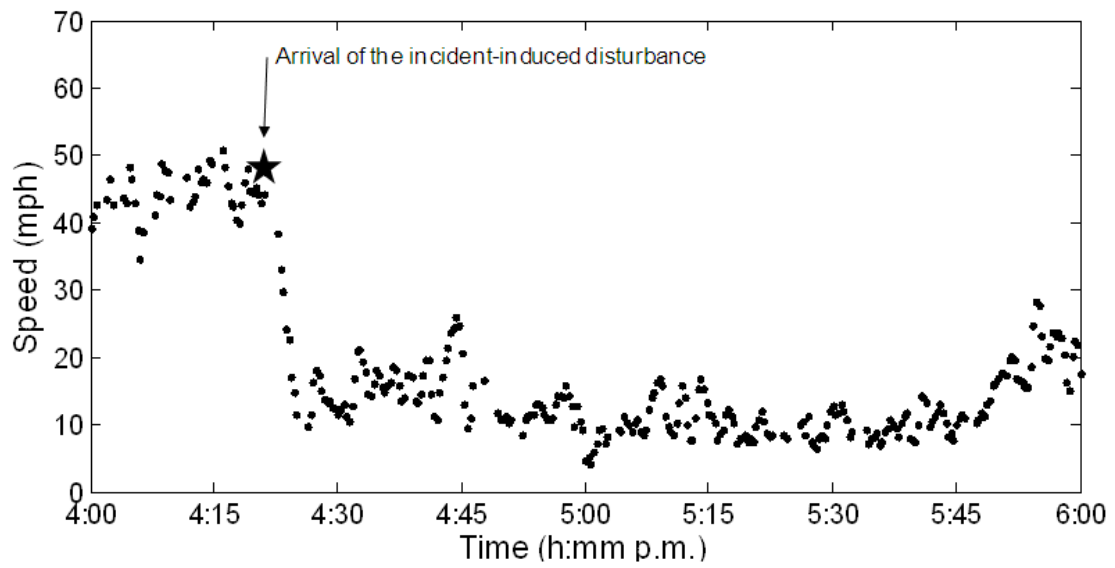
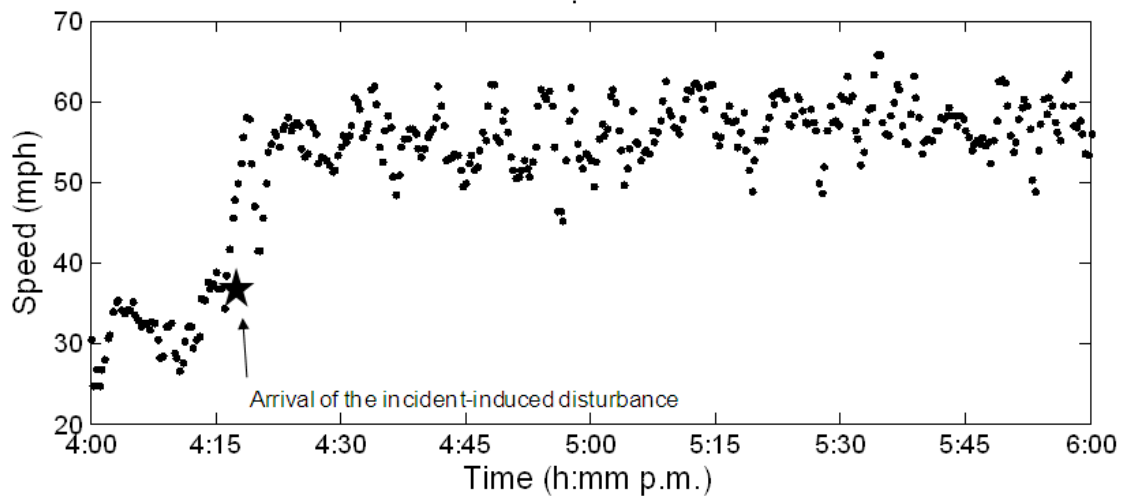


Figure 1.4 Data quality map of northbound I-5 during p.m. peak hours

Figure 1.5 demonstrates this assumption using time-series speeds at the detector stations immediately upstream (Macadam Ave.) and downstream (Morrison Bridge) of a crash that occurred at milepost 300.31. According to the crash database, this crash occurred sometime between 4 and 5 p.m. on April 26, 2006. The time-series speed at the upstream detector station (Figure 1.5(a)) initially displays relatively moderate congestion with a speed of 45 mph (73 km/h). However, around 4:20 p.m. (marked by a star in the figure), the speed decreases markedly to around 10-15 mph (16-25 km/h), marking the arrival of the disturbance due to the crash downstream. The opposite trend is observed at the downstream location (Figure 1.5(b)), which is initially characterized by a congested speed of about 30 mph (49 km/h). However, around 4:18 p.m., the speed increases to about 55-60 mph (89-97 km/h), marking the arrival of the crash-induced disturbance. Since this crash occurred in the influence area of the loop station at Macadam Ave. (see Table 1), data from this upstream station taken prior to 4:20 p.m. are used to represent pre-crash traffic conditions. Crashes whose occurrence times could not be determined from the incident database or traffic data were excluded from the analysis. (Traffic data from the days of these crashes were also excluded.) A total of 194 crashes remained (out of 242).



(a)



(b)

Figure 1.5 Illustration of crash occurrence time estimation

The remaining 194 crash events were further scrutinized and removed if they met any of the following three criteria: (1) the crash occurred in an uncongested condition; (2) traffic data before the crash are invalid; or (3) an involved driver was impaired by

drugs/alcohol. After these filtering processes, a total of 82 crashes remained; these were used in the analysis.

The majority of the 82 crashes were rear-end crashes (71 out of 82), which seems reasonable. Nine sideswipe-overtaking crashes (vehicles traveling in the same direction on parallel paths collide) and one fixed or other object crash (a vehicle strikes a fixed or other object on the roadway or off roadway) were observed. One crash was coded as a turning movement crash, but inspection of the entire crash record indicates that this was probably mis-coded. The crash occurred on the mainline in the vicinity of an exit ramp. Both vehicles were traveling in the same direction and likely collided when a vehicle made a maneuver towards the exit ramp (the reason for the turn code). As all these crashes can occur due to oscillations, all 82 crashes were included in the analysis. (The descriptions of the turning-movement and fixed or other object crashes were not clear enough to determine if the crashes were potentially induced by oscillations. To be on the conservative side, these crashes were included in the analysis. Nevertheless, these crashes are unlikely to affect the results significantly since there were only two crashes of these crash types.)

Finally, the 82 crash events were linked to traffic data based on their mileposts and the corresponding influence areas. Of note, this study used 1-minute moving averages of traffic measures (e.g., speed) to smooth out noise while preserving the underlying trend. Measurements from all lanes were also combined to incorporate influences from adjacent lanes and maximize the sample size because Abdel-Aty et al. (2005) pointed out that a sample size can be maximized by using measurements taken in other lanes in case of a

malfunctioning loop. Weather information (e.g., rainfall) was obtained from PORTAL (2009) and incorporated into the crash and traffic data.

To verify the assumption of uniform oscillations within each influence area, oscillations were sampled during a 10-minute period (around 5 p.m.) each day at the most downstream location and measured the amplitude of oscillations by taking the standard deviation of speeds. Note that the duration of 10 minutes was used since oscillations typically exhibited a comparative period. This is further elaborated in the following section. The amplitudes of the sampled oscillations were then traced at all upstream locations by accounting for the oscillation propagation speed. As noted in Mauch and Cassidy (2002) and Ahn and Cassidy (2007), oscillations in congestion propagate upstream against traffic flow at nearly constant speeds independent of traffic states. The propagation speed from one detector station to the immediately upstream station was estimated using the cross-correlation technique. The basic idea is to search for a time lag which maximizes the correlation coefficient between the time series from two detector stations. The optimal time lag corresponds to the travel time of oscillations between the two locations. The propagation speed corresponding to the optimal time lag was found to be around 12 mph.

Figure 1.6 shows the basic statistics (e.g., average and standard deviations) of oscillations amplitude at each detector station for all crash-free days, as well as the spatial distribution of the identified crashes. The bars stand for number of crashes; the solid line is for average oscillations and the dotted lines are variations of oscillations over time bounded by one standard deviation. The figure shows that the average amplitudes of

oscillations are reasonably stable over space, such that they did not deviate more than 0.4 mph within each influence area on average. The largest difference of 0.8 mph is observed at the Multnomah station based on linearly interpolated amplitudes, though only 4 crashes occurred within this influence area. Thus, it seems reasonable to assume that oscillations within an influence area can be reasonably approximated by the oscillations measured at the corresponding loop detector station.

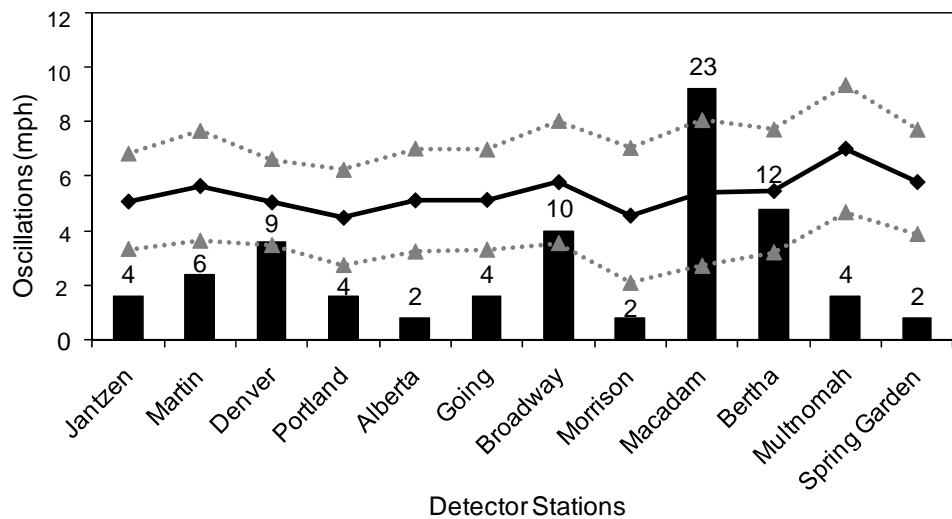


Figure 1.6 Spatial distributions of oscillations and number of crashes sampled

1.4 Methodology: Case Control Design

The matched case-control method was selected to study the relationship between traffic oscillations and crash occurrences in congestion. Many previous studies (mentioned in the Background section) found that crash occurrences are significantly associated with exogenous factors such as traffic conditions (e.g. speed), roadway characteristics (e.g. pavement conditions, speed limits, geometric features), and environmental conditions (e.g., weather). Thus, these factors have been incorporated in

the modeling. Since the primary objective of this study is to quantify the impact of traffic operational features, the factors related to roadway and environmental conditions are controlled in the matched case-control design.

Two types of exposure variables are considered for traffic conditions, one for average traffic states and one for oscillations during a period of time (10 minutes, as explained below) prior to the crashes. Potential measures of oscillations include the standard deviations of speed, count, and occupancy. The mean speed, count and occupancy are used as measures of average traffic states.

The measurement duration of the explanatory variables was determined based on their temporal characteristics: an ideal duration would be long enough to capture at least one oscillation cycle but short enough to preclude changes in longer-term average traffic states. The ideal sampling duration was determined using oblique curves of cumulative time-mean speed. Time-series plots of count, occupancy and speed typically display a large amount of noise, making it difficult to identify different traffic states and the times of state changes. To smooth the noise in the data and better reveal the underlying trends, cumulative curves of traffic observations are constructed on an oblique time axis. This technique is described in detail in Munoz and Daganzo (2002). On an orthogonal axis, an oblique curve is obtained by taking the difference between the cumulative speed at time t , $V(t)$, and its background reduction, $V_0 * (t-t_0)$, where V_0 is a scaling factor and t_0 is the starting time of the curve. Figure 1.7 presents an oblique curve of cumulative speed, $V(x, t) - V_0 * (t-t_0)$, versus time, t , at Jantzen Dr. between 4:30 and 5:30 p.m. on May 14,

2007. The figure shows a series of oscillations with a cyclic pattern of an increase in slope (i.e., speed) followed by a decrease.

To get some basic idea on the duration of typical oscillations, short time fourier transform (STFT) is implemented for the speed data at Jantzen Dr. between 4:30 and 5:30 p.m. on May 14, 2007. Power spectral density (PSD) estimated by STFT is shown as in Figure 1.8. PSD is a function of frequency and measures amount of power (energy) per unit (density) of frequency. In this figure, the brightness represents the energy contained in a specific frequency. Judging from the figure, most of energy concentrates on a frequency of around 0.0016 Hz (the dash-dotted line in the figure), which corresponds to 10 minutes per cycle. Therefore, oscillations typically have a period of about 10 minutes; this was used as the sampling duration for this study.

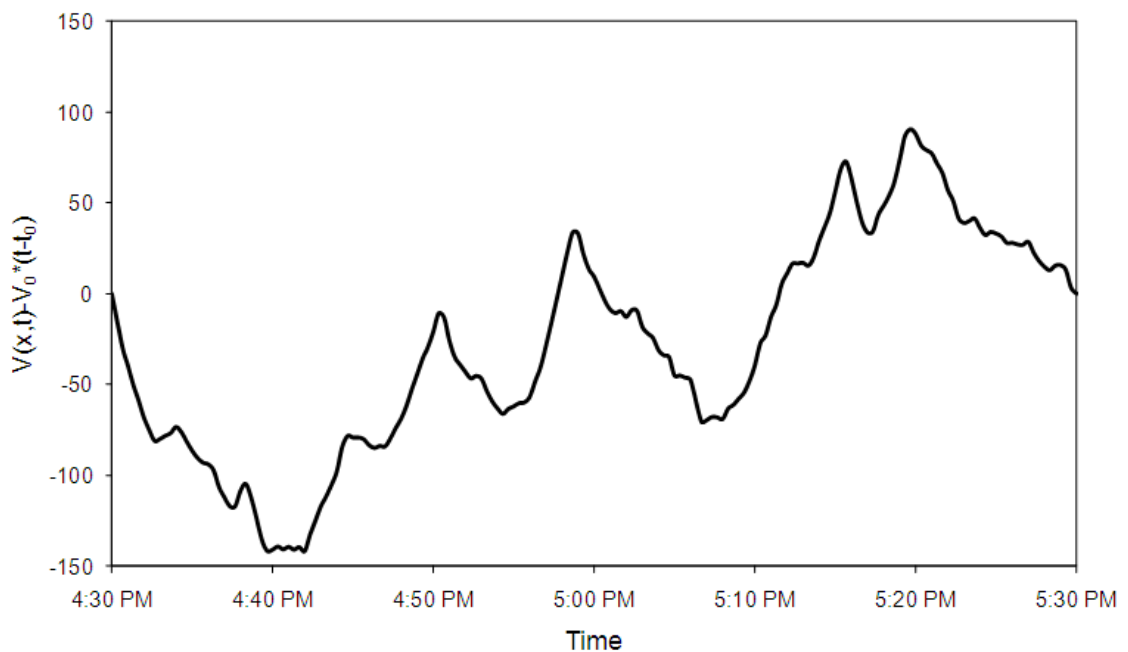


Figure 1.7 Oblique curve of cumulative speed at Jantzen Dr.

(between 4:30 and 5:30 p.m. on May 14, 2007)

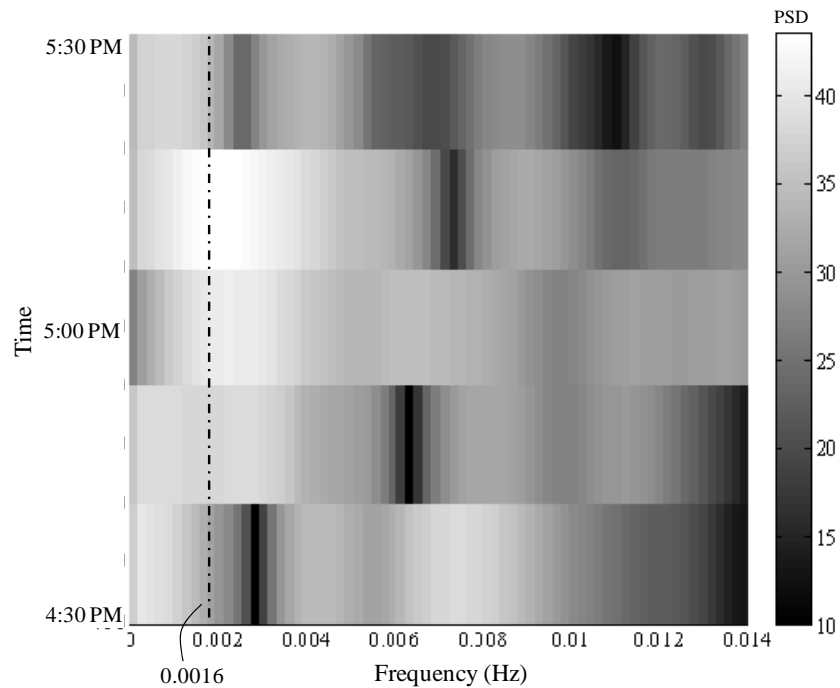


Figure 1.8 STFT plot of the speed data at Jantzen Dr.

(between 4:30 and 5:30 p.m. on May 14, 2007)

In the literature, it has been reported that crash occurrence times are often rounded to the nearest 5 minutes (Golob and Recker 2003; Kockelman and Ma 2007). In such cases, traffic data within the 5 minutes before each crash should be excluded from the analysis to avoid “cause and effect” ambiguity. The dataset was inspected to assess if the crash occurrence times were rounded to the nearest 5 minutes in Oregon. The analysis shows that the majority (68 out of 82) of the crash occurrence times are not in multiples of 5 minutes. Therefore, the author believe that the crash occurrence times used in this

research were not rounded to the nearest 5 minutes and thus, did not further exclude any data.

The dependent variable is binary (crash-prone or not-crash-prone traffic conditions). Of note, the cases here are not the crashes themselves, but the traffic conditions immediately before the crash occurrences, while the controls are congested traffic conditions under which no crash has occurred.

To control for the potential impacts of other factors on crash occurrences, several confounding variables, time of day, location, weather, and presence of congestion, are used to limit the selection of controls. Of note, each location is treated as a unique geometric factor to be conservative. For instance, a crash occurred in the influence area of the Broadway St. detector station at 4:11 p.m. on the rainy day of October 6, 2005. Measurements taken from 4:01 to 4:11 p.m. on this day are used as a case. Then, measurements are taken at the same location during the same period on the crash-free days in 2004 to 2007 that exhibited congestion (average speed less than 50 mph (81 km/h)) and similar weather conditions. These extracted measurements are used as potential candidates for controls. Among the candidates, n observations are randomly selected as the controls for that particular crash (case). In this study, about 1,000 candidate controls are available for each case. All legitimate controls were stored for the model evaluation, as discussed in Section 6.

The above procedure is applied to each crash, resulting in the 82 cases matched with $82 \times n$ controls. It is difficult to theoretically determine the optimal control-to-case ratio. As a rule of thumb, a control-to-case ratio around 4:1 is recommended since the statistical

power generally does not increase significantly beyond a 4:1 ratio (Ghosh and Chen 2002; Hennekens and Buring 1987; Schlesselman and Stolley 1982). Therefore, the control-to-case ratio of 4:1 was implemented in this study.

The modeling efforts essentially consist of two main phases: 1) model development using a set of sampled controls and 2) model evaluation via repeated model developments using different samples of controls and sensitivity analysis with respect to the control-to-case ratio (4:1 to 7:1). The latter is designed to evaluate the consistency of modeling results with respect to multiple (unique) control samples and control-to-case ratios. These efforts are presented in detail in Section 5 and Section 6.

1.5 Model Development

Conditional logistic regression is employed to model the association between traffic oscillations and crash occurrences in the study segment. This is a standard technique for analyzing matched case-control data. Conditional inference can be adopted easily in the conventional (unconditional) logistic regression model if the constant for each stratum is excluded in the estimation by treating it as a “nuisance” parameter (Hosmer and Lemeshow 2004). Estimates from a conditional logistic regression have been proven to be consistent and asymptotically normally distributed. Cox and Hinkley (1974) provide the mathematical details of conditional likelihood analysis.

Conditional logistic regression analysis is implemented in Stata[®]/IC10 (StataCorp 2009) by using average and the standard deviations of flow, density or speed as potential explanatory variables of average traffic states and oscillations. It is well acknowledged in the literature that the control-to-case ratio should be around 4:1, as cited in the previous

section, since 80 percent of the maximum efficiency can be obtained around that ratio.

Table 1.2 lists all the potential explanatory variables and their basic statistics. In this table, the average speed is calculated across all travel lanes at all detector stations during the ten-minute periods right before the crash occurrences; the average count is the average number of vehicles per lane per 20 seconds at all detector stations during the ten-minute periods right before the crash occurrences; the average occupancy is the average percentage of time that the detectors are occupied by vehicles across all travel lanes during the ten-minute periods right before the crash occurrences; Std. dev. of speed is the standard deviation of speed during the ten-minute periods right before the crash occurrences; and std. dev. of count and std. dev. of occupancy are similarly defined. The average speed is about 25 mph (40 km/h), indicating heavy congestion. The mean amplitude of traffic oscillations in terms of the standard deviation of speed is 5 mph (8 km/h). The standard deviation of oscillations (also in terms of standard deviation of speed) is 3 mph (5 km/h), but ranges from 0.3 mph (0.5 km/h) to nearly 20 mph (32 km/h), confirming that the samples (a total of 82 cases and 328 controls) contain variations in traffic oscillations.

Table 1.2

Basic statistics of the potential explanatory variables

(4:1 control-to-case ratio)

Potential Variable	Mean	Std. Dev.	Min	25% Percentile	75% Percentile	Max
Average speed	24.743	11.237	2.303	16.542	30.448	49.938
Average count	5.294	1.656	0.703	4.072	6.444	9.559
Average occupancy	20.955	14.913	0.486	4.968	33.401	64.294
Std. dev. of speed	4.982	3.057	0.327	2.928	6.015	19.957
Std. dev. of count	0.894	0.402	0.0993	0.595	1.098	2.460
Std. dev. of occupancy	4.364	3.829	0.380	1.233	6.386	30.609

A sound model development practice was followed. Namely, different combinations of potential explanatory factors were incorporated via conditional logistic regression, following a univariate analysis. However, since flow, density and speed are highly correlated in congested traffic (e.g., transitioning into a more congested state is typically accompanied by decreases in flow and speed and an increase in occupancy), only one of them is used to represent average traffic state or oscillations in the modeling process. Therefore, two exposure variables (one for oscillations and one for average traffic state) are included. Furthermore, as supported by previous studies, an interaction term is not considered for the lack of a strong theoretical insight.

There are many criticisms (Good and Hardin 2006; Johnson 1999; Sterne and Smith 2001) of over-emphasizing significance tests, such as the p-values of individual parameters and (Pseudo) R^2 , to evaluate models, as these tests can be arbitrarily manipulated (by changing the sample size or adding unnecessary variables, for example) and are difficult to interpret correctly. Thus, the performance of each model was evaluated based on the theoretical soundness (e.g., signs of estimated coefficients), the p -

value for the overall model, R^2 and likelihood ratio tests, rather than the statistical significance of each estimated parameter. Of note, the threshold p -value for the overall model performance is set to 0.1; models with overall p -values less than 0.1 were considered as candidates for the best model. Finally, a model with the best performance was selected.

Table 1.3 shows the statistics of the best model for a set of sampled controls with 4:1 control-to-case ratio. The result shows that the odds ratio for the standard deviation of speed is about 1.096, with a 95 percent confidence interval of 1.016 to 1.183. This implies that if the standard deviation of speed increases by one additional unit (1 mph in this case), the odds ratio of a crash occurrence would increase by 9.6 percent.

Table 1.3

Results from the conditional logistic regression analysis

(4:1 control-to-case ratio)

Crash_prone	Odds Ratio	Std. Err.	Z	P> z	95% Confidence Interval (CI)	
Std. dev. of speed	1.096	.0424	2.37	0.02	1.016	1.183
Number of observations=410; LR chi2(1) = 5.37; Prob > chi2 =0.0205; Log likelihood = -129.29						

Since the independent variables in the model are continuous, the linearity assumption in the logistic model has been tested using fractional polynomials (FP), as recommended by Hosmer and Lemeshow (2004). The fractional polynomials allow us to build several models with different power functions of a continuous explanatory variable and then to choose the most suitable one by conducting the partial likelihood ratio test. The analysis indicates that keeping the continuous variables linear in the logistic model is appropriate

for the range of data (e.g., speeds). The details of this analysis are provided in Appendix A.

Several techniques (e.g., computing residual variation, leverage) can be used to detect potential outliers in a conditional logistic regression (Hosmer and Lemeshow 2004). However, removing outliers potentially can lead to over-fitting, which diminishes the effectiveness of the model (Washington et al. 2003). Therefore, all data points are included in the modeling efforts and evaluate the effectiveness of the model using different samples of controls, as described in the following section.

1.6 Model Evaluation and Interpretation

Evaluation of statistical models for consistency prior to interpretation is essential since they are often subject to measurement errors, selection bias, invalid design, and/or human errors. Ideally, cases and controls should be free of any sources of bias, such as construction, lane closure, and other incidents. This issue was not directly addressed in the analysis given the extensive data collection efforts involved. Instead, the model was evaluated by re-sampling controls for each case and repeating the model development process (described in Section 2.5) to find the best models based on the newly drawn samples and conducting a sensitivity analysis with respect to the case-to-control ratio. For the former, twenty different sets of controls were sampled from the pool of candidate controls for the control-to-case ratio 4:1. Note that sampling the same control more than once is unlikely since each case has about 1,000 candidate controls. The best model for each re-sampled dataset was obtained according to the criteria described in Section 2.5.

Table 1.4 presents the modeling results from the 20 different runs for the control-to-case ratio of 4:1. In this table, avg_occ is average occupancy; std_spd is standard deviation of speed; avg_cnt is average count, and N/A means not available.

The result indicates that in 18 out of 20 runs (90 percent), the standard deviation of speed is significant, and its odds ratio is fairly consistent throughout the runs. Notably, the average odds ratio is 1.084 with a minimum of 1.066 and maximum of 1.119, which is fairly a tight bound. Moreover, the average lower and upper limits of the 95 percent confidence intervals are 1.003 and 1.171, respectively. For the study segment, an additional unit increase in the standard deviation of speed increases the odds ratio of crash occurrence by an average of 8.4 percent. This indicates that an average day is about 1.49 times (calculated as $(1+0.084)^{4.982} = 1.49$) more likely to have a crash than the day without any oscillations since the average standard deviation of speed on the study corridor is about 4.982 mph (see Table 1.2). This magnitude of impact seems reasonable and quite significant since many other factors, such as human factors and vehicle conditions, also impact crash occurrences (Hauer 1997).

Table 1.4

Model evaluation results (4:1 control-to-case ratio)

Run	Variables	Odds Ratio	95% CI
1	std_spd	1.096	(1.016, 1.183)
2	std_spd	1.066	(0.988, 1.151)

3	avg_occ	1.030	(1.002, 1.058)
	std_spd	1.119	(1.029, 1.216)
4	std_spd	1.070	(0.988, 1.160)
5	std_spd	1.090	(1.010, 1.177)
6	std_spd	1.069	(0.994, 1.149)
7	std_spd	1.077	(1.001, 1.160)
8	std_spd	1.079	(1.006, 1.158)
9	avg_occ	1.020	(0.997, 1.043)
	std_spd	1.071	(0.991, 1.157)
10	avg_occ	1.025	(0.997, 1.054)
	std_spd	1.094	(1.010, 1.185)
11	avg_occ	1.028	(0.999, 1.058)
	std_spd	1.118	(1.029, 1.215)
12	avg_occ	1.019	(0.995, 1.042)
	std_spd	1.089	(1.008, 1.176)
13	avg_occ	1.020	(0.996, 1.044)
	std_spd	1.077	(0.995, 1.165)
14	std_spd	1.081	(1.002, 1.166)
15	N/A	N/A	N/A
16	avg_cnt	0.809	(0.646, 1.013)
	std_spd	1.089	(1.005, 1.179)
17	avg_occ	1.017	(0.994, 1.041)
	std_spd	1.069	(0.994, 1.150)
18	avg_occ	1.021	(0.994, 1.050)
	std_spd	1.081	(0.997, 1.172)
19	avg_occ	1.026	(0.999, 1.053)
20	std_spd	1.078	(0.998, 1.165)
Average	std_spd	1.084	(1.003, 1.171)

It is also notable that the average occupancy is significant only 9 times in the 20 runs. Thus, it is reasonable to conclude that speed variations in queued traffic (i.e., oscillations) have a larger impact on crash occurrence than average traffic states. Nevertheless, the average odds ratio for occupancy (around 1.02) is qualitatively consistent: it implies that the likelihood of crash occurrence increases as congestion becomes more severe.

Now the author turns attention to the sensitivity analysis of the modeling results shown in Table 1.4 with increasing control-to-case ratios. If the model results are robust, we expect similar or more consistent results for larger control-to-case ratios since the

statistical power is expected to increase. To verify this, additional controls were sampled and added to the controls for the 4:1 ratio; for example, for the 5:1 ratio, one additional control for each case was randomly sampled from the pool of candidates and added to the existing four controls. For each run and each increased control-to-case ratio, a best model was re-developed based on the criteria described in Section 5 to examine the consistency of models in terms of significant explanatory variables, odds ratios, and confidence intervals. The results of this sensitivity analysis are summarized in Table 1.5 for the control-to-case ratios of 4:1 to 7:1. In this table, avg_occ is average occupancy; std_spd is standard deviation of speed; avg_cnt is average count, and N/A means not available.

Table 1.5 reports the number of runs in which the standard deviation of speed (i.e., amplitude of oscillations) and average occupancy, speed, and count were significant over 20 different control samples for each control-to-case ratio. The table also reports the average odds ratios and average confidence intervals for the standard deviation of speed. The model results for the standard deviation of speed are quite consistent for different control samples, and the consistency improves with increasing ratios, as expected. More specifically, the standard deviation of speed is significant in more than 18 runs for all ratios, and the number of significant runs increases from 18 (90 percent of the runs) to 20 (100 percent of the runs) as the ratio increases from 4:1 to 7:1. Moreover, the average odds ratios for the standard deviation of speed are consistently around 1.08 for different control-to-case ratios, and the average 95 percent confidence intervals tighten as the ratio increases. Thus, the results strongly support the existence of relation between oscillations (measured by the standard deviation of speed) and crash occurrences. To the contrary,

other variables for average traffic states appeared to be significant in less than 10 runs, and the consistency does not improve with increasing control-to-case ratios. The results imply that traffic oscillations are more significantly associated with crash occurrences than average traffic states. The details of individual runs for each control-to-case ratio are presented in Table B1 in Appendix B.

Table 1.5

Model evaluation results for different control-to-case ratios

(4:1 to 7:1)

Control-to-Case Ratio	Standard Deviation of Speed			Number of Significant Runs out of 20			
	Average Odds Ratio	Average 95% CI		std_spd	avg_occ	avg_spd	avg_cnt
4:1	1.084	1.003	1.171	18	9	0	1
5:1	1.080	1.002	1.165	18	4	1	1
6:1	1.080	1.003	1.163	19	8	1	1
7:1	1.081	1.005	1.162	20	5	3	0

REFERENCES

- Abdel-Aty, M., Pande, A., Uddin, N., Dilmore, J., Pemmanaboina, R. (2005). *Relating crash occurrence to freeway loop detectors data, Weather Conditions and Geometric Factors*. University of Central Florida.
- Addison, P. S. (2002). *The illustrated wavelet transform handbook: Introductory theory and applications in science, engineering, medicine and finance*, Taylor & Francis.
- Ahn, S. and Cassidy, M. J. (2007). Freeway traffic oscillations and vehicle lane-change maneuvers. *Proc., 17th International Symposium of Transportation and Traffic Theory*, 691-710, Amsterdam, Elsevier.
- Ahn, S., Laval, J., Cassidy, M. (Accepted , 2010). Merging and diverging effects on freeway traffic oscillations: theory and observations. *Transportation Research Record, Journal of the Transportation Research Board*.
- Ahn, S. (2005). *Formation and spatial evolution of traffic oscillations*. Doctoral Dissertation, Department of Civil and Environment Engineering, University of California, Berkeley.
- Ahn, S., Cassidy, M. J., Laval, J. (2004). Verification of simplified car-following theory. *Transportation Research Part B*, 38(5): 431-440.
- Banks, J. (1991). The two-capacity phenomenon: some theoretical issues. *Transportation Res. Record*, 1320: 234-241.
- Bilbao-Ubillo, J. (2008). The costs of urban congestion: Estimation of welfare losses arising from congestion on cross-town link roads. *Transportation Research Part A*, 42(8), 1098-1108.
- Brackstone, M., McDonald, M. (1999). Car-following: A historical review. *Transportation Research Part F*, 2(4), 181-196.
- Breslow, N. E., Day, N. E. (1980). Statistical methods in cancer research. Volume I - the analysis of case-control studies. *IARC Sci. Publ.*, 32, 5-338.
- Briggs, W. M., Levine, R. A. (1997). Wavelets and field forecast verification. *Mon. Weather Rev.*, 125(6), 1329-1341.
- Brockfeld, E., Kühne, R. D., Skabardonis, A., Wagner, P. (2002). Toward a benchmarking of microscopic traffic flow models. *Transportation Research Record: Journal of the Transportation Research Board*, 1852: 124-129.

- Cassidy, M. J., and Rudjanakanoknad, J. (2005). Increasing the capacity of an isolated merge by metering its on-ramp. *Transportation Research Part B*, 39(10), 896-913.
- Cassidy, M. J., and Bertini, R. L. (1999). Some traffic features at freeway bottlenecks. *Transportation Research Part B*, 33(1), 25-42.
- Cassidy, M., and Windover, J. (1995). Methodology for assessing dynamics of freeway traffic flow. *Transp. Res. Rec.*, (1484), 73-79.
- Castillo, J. M. (2001). Propagation of perturbations in dense traffic flow: A model and its implications. *Transportation Research Part B*, 35(4), 367-389.
- Ceder, A., and May, Jr., A. D. (1976). Further evaluation of single and two regime traffic flow models. *Transportation Research Record*, 567, 1-30.
- Chandler, R. E., Herman, R., Montroll, E. W. (1958). Traffic dynamics: Studies in CF. *Oper. Res.*, 6(2), 165-184.
- Chiabaut, N., Buisson, C., Leclercq, L. (2009). Fundamental diagram estimation through passing rate measurements in congestion. *IEEE Transactions on Intelligent Transportation Systems*. 10(2): 355-359.
- Cirillo JA. (1968). Interstate system accident research: Study II, Interim Report II. 35(3), 71-75.
- Coifman, B., Mishalani, R., Wang, C., Krishnamurthy, S. (2006). Impact of lane-change maneuvers on congested freeway segment delays: Pilot study. *Transp. Res. Rec.*, 1965, 152-159.
- Coifman, B., Krishnamurthy, S., Wang, X. (2005). Lane-change maneuvers consuming freeway capacity, In *Traffic and Granular Flow '03*, 3-14, Berlin, Springer.
- Cornfield, J., Haenszel, W., Hammond, E. C., Lilienfeld, A. M., Shimkin, M. B., Wynder, E. L. (1959). Smoking and lung cancer: Recent evidence and a discussion of some questions. *J. Natl. Cancer Inst.*, 22(1), 173-203.
- Cornfield, J. (1951). A method of estimating comparative rates from clinical data; Applications to cancer of the lung, breast, and cervix. *J. Natl. Cancer Inst.*, 11(6), 1269-1275.
- Cox, D. R., and Hinkley, D. (1974). *Theoretical statistics*, Chapman & Hall/CRC.
- Daganzo, C.F. (1995). Requiem for second-order fluid approximations of traffic flow. *Transportation Research Part B*, 29, 277-286.

- Daganzo, C. F. (2002). A behavioral theory of multi-lane traffic flow. Part I: Long homogeneous freeway sections. *Transportation Research Part B*, 36(2), 131-158.
- Daganzo, C.F. (2005). A variational formulation of kinematic waves: basic theory and complex boundary conditions. *Transportation Research Part B*, 39(2), 187-196.
- Daubechies, I. (1992). *Ten lectures on wavelets*, Society for Industrial and Applied Mathematics (SIAM).
- Davis, G. A., Davuluri, S., Pei, J. (2006). Speed as a risk factor in serious run-off-road crashes: Bayesian case-control analysis with case speed uncertainty. *Journal of Transportation and Statistics*, 9(1), 17.
- Davis, G. A. (2002). Is the claim that 'variance kills' an ecological fallacy? *Accid. Anal. Prev.*, 34(3), 343-346.
- Donoho, D. L. (1993). Unconditional bases are optimal bases for data compression and for statistical estimation. *Applied and Computational Harmonic Analysis*, 1(1), 100-115.
- Duret, A., Buisson, C., Chiabaut, N. (2007). Estimating individual speed-spacing relationship and assessing ability of Newell's car-following model to reproduce trajectories. *Transportation Research Record: Journal of the Transportation Research Board*. 2088, 188-197.
- Duret, A., Ahn, S., Buisson, C. (Accepted with revision, 2010). Passing rates to measure relaxation and impact of lane-changing in queue. *Computer Aided Engineering and Infrastructure Engineering*.
- Edie, L. C. (1960). Car following and steady state theory for non-congested traffic. *Operations Research*, 9, 66-76.
- Edie, L. C., and Baverez, E. (1965). Generation and propagation of stop-start traffic waves. *Vehicular Traffic Science, Proc. 3rd Internat. Sympos. Theory Traffic Flow*, New York, 26-37.
- Edie, L. C. (1963). Discussion of traffic stream measurements and definitions, *Proc. 2nd Internat. Sympos. Theory Traffic Flow*, 139-154. London.
- Edie, L. C. (1961). Car-following and steady-state theory for non-congested traffic. *Oper. Res.*, 9(1), 66-76.
- Farge, M. (1992). Wavelet transforms and their applications to turbulence. *Annu. Rev. Fluid Mech.*, 24(1), 395-458.
- FHWA. The next generation simulation (NGSIM). <<http://www.ngsim.fhwa.dot.gov/>> (05/08, 2008).

- Fildes, B., and Rumbold, G. (1991). *Speed behavior and drivers' attitude to speeding*, Monash University.
- Fisher, R. A. (1958a). Lung cancer and cigarettes. *Nature*, 182(4628), 108.
- Fisher, R. A. (1958b). Cancer and smoking. *Nature*, 182(4635), 596.
- Fowles, R., and Loeb, P. D. (1989). Speeding, coordination, and the 55-mph limit: comment. *Am. Econ. Rev.*, 79(4), 916-921.
- Fritzsche, H. (1994). A model for traffic simulation. *Traffic Engineering and Control*, 35(5), 317-321.
- Gazis, D. C. (2002). The origins of traffic theory. *Oper. Res.*, 50(1), 69-77.
- Gazis, D. C., Herman, R., Rothery, R. W. (1961). Nonlinear follow-the-leader models of traffic flow. *Oper. Res.*, 9(4), 545-567.
- Gazis, D. C., Herman, R., Potts, R. B. (1959). Car-following theory of steady-state traffic flow. *Oper. Res.*, 7(4), 499-505.
- Ghosh, M., and Chen, M. H. (2002). Bayesian inference for matched case-control studies. *Sankhyā: The Indian Journal of Statistics, Series B*, 107-127.
- Gipps, P. (1981). A behavioral car-following model for computer simulation. *Transportation Research Part B*, 15(2), 105-111.
- Golob, T. F., and Recker, W. W. (2003). Relationships among urban freeway accidents, traffic flow, weather, and lighting conditions. *J. Transp. Eng.*, 129-342.
- Good, P. I., and Hardin, J. W. (2006). *Common errors in statistics (and how to avoid them)*, Wiley-Interscience.
- Greenwood, I. D., and Bennett, C. R. (1996). The effects of traffic congestion on fuel consumption. *Road and Transport Research*, 5, 18-35.
- Hauer, E. (1997). *Observational before-after studies in road safety*. New York, Elsevier.
- Helbing, D. (1997). Fundamentals of traffic flow. *Physical Review E*, 55(3), 3735-3738.
- Helbing, D. (1998). Structure and instability of high-density equations for traffic flow. *Phys. Rev. E*, 57, 6176-6179.
- Hennekens, C. H., and Buring, J. E. (1987). *Epidemiology in medicine*, Lippincott Williams & Wilkins.

- Herman, R., Montroll, E. W., Potts, R. B., Rothery, R. W. (1959). Traffic dynamics: analysis of stability in car following. *Oper. Res.*, 7, 86-106.
- Heyes, M. P., and Ashworth, R. (1972). Further research on car following models. *Transportation Research*, 6, 287-291.
- Hosmer, D. W., and Lemeshow, S. (2004). *Applied logistic regression*, Wiley-Interscience.
- Ioannou, P., and Stefanovic, M. (2005). Evaluation of ACC vehicles in mixed traffic: Lane change effects and sensitivity analysis. *IEEE Transactions on Intelligent Transportation Systems*, 6(1), 79-89.
- Johnson, D. H. (1999). The insignificance of statistical significance testing. *The Journal of Wildlife Management*, 763-772.
- Katul, G., and Vidakovic, B. (1996). The partitioning of attached and detached eddy motion in the atmospheric surface layer using Lorentz wavelet filtering." *Bound.-Layer Meteorol.*, 77(2), 153-172.
- Kerner, B. (1998). Experimental features of self-organization in traffic flow. *Phys. Rev. Lett.*, 81(17), 3797-3800.
- Kerner, B., and Rehborn, H. (1996). Experimental features and characteristics of traffic jams. *Physical Review E*, 53(2), 1297-1300.
- Kim, T., and Zhang, H. (2008). A stochastic wave propagation model. *Transportation Research Part B*, 42, 619-634.
- Kloeden, C., A. McLean, V. Moore, G. Ponte. (1997). *Travelling speed and the risk of crash involvement*, NHMRC Road Accident Research Unit, University of Adelaide, Adelaide, Australia.
- Kockelman, K. M., and Ma, J. (2007). Freeway speeds and speed variations preceding crashes, within and across lanes. *Journal of the Transportation Research Forum*, 46(1), 43-62.
- Kometani, E., and Sasaki, T. (1959). Dynamic behavior of traffic with a nonlinear spacing-speed relationship. In *Proceedings of the Symposium on Theory of Traffic Flow*, Research Laboratories, General Motors, 105-119, New York, Elsevier.
- Koshi, M., Iwasaki, M., Ohkura, I. (1983). Some findings and an overview on vehicular flow characteristics. *Proc., Proceedings of the 8th International Symposium on Transportation and Traffic Theory*, 403-426, University of Toronto Press, Toronto, Canada.

- Laval, J.A., and Leclercq, L. (Forthcoming, 2010). A mechanism to describe the formation and propagation of stop-and-go waves in congested freeway traffic. *Proceedings of The Royal Society A*.
- Laval, J. A., and Leclercq, L. (2008). Microscopic modeling of the relaxation phenomenon using a macroscopic lane-changing model. *Transportation Research Part B*, 42, 511-522.
- Laval, J. A., and Daganzo, C. F. (2006). Lane-changing in traffic streams. *Transportation Research Part B: Methodological*, 40(3), 251-264.
- Lave, C. (1989). Speeding, coordination, and the 55 mph limit: Reply. *Am. Econ. Rev.*, 79(4), 926-931.
- Lave, C. A. (1985). Speeding, coordination, and the 55 mph Limit. *Am. Econ. Rev.*, 75(5), 1159-1164.
- Leclercq, L., Chiabaut, N., Laval, J., Buisson, C. (2007). Relaxation phenomenon after changing lanes: Experimental validation with NGSIM data set. *Transportation Research Record*, 1999:79-85.
- Levy, D. T., and Asch, P. (1989). Speeding, coordination, and the 55 mph limit: Comment. *Am. Econ. Rev.*, 79(4), 913-915.
- Lighthill, M., and Whitham, G. (1955). On kinematic waves: II. A theory of traffic flow on long crowded roads. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 229(1178), 317-345.
- Lownes, N. E., and Machemehl, R. B. (2006). VISSIM: A multi-parameter sensitivity analysis. *Proc., Proceedings of the 38th Conference on Winter Simulation Conference*, 1406-1413.
- Ma, T., and Ahn, S. (2008). Comparisons of speed-spacing relations under general CF versus lane changing. *Transportation Research Record, Journal of the Transportation Research Board*, 2088, 138-147.
- Mallat, S. G. (1999). *A wavelet tour of signal processing*, Academic Press.
- Malyshkina, N., and Mannering, F. L. (2008). Analysis of effect of speed-limit increases on accident causation and injury severity. *Proc., TRB 87th Annual Meeting Compendium of Papers DVD*, Washington, DC.
- Manski, C. F. (1995). *Identification problems in the social sciences*, Harvard University Press.

- Mauch, M., and Cassidy, M. J. (2002). Freeway traffic oscillations: Observations and predictions. *Proc., 15th Int. Symp. on Transportation and Traffic Theory*, 653–674. Pergamon-Elsevier, Oxford, UK.
- Mauch, M. (2002). *Analyses of start-stop waves in congested freeway traffic*. Doctoral Dissertation, Department of Civil and Environmental Engineering, University of California, Berkeley.
- May, Jr., A. D, and Keller, H. E. M. (1967). Non integer car following models. *Highway Research Record*, 199, 19-32.
- Meyer, Y., and Salinger, D. (1995). *Wavelets and operators*, Cambridge University Press.
- Michaels, R. (1965). Perceptual factors in car following. *Proc. of the 2nd Int. Symp on the Theory of Road Traffic Flow*, 44-59.
- Mika, H., Kreer, J., Yuan, L. (1969). Dual mode behavior of freeway traffic. *Highway Research Board 41th annual meeting*, Washington, DC.
- Muñoz, J. C., Daganzo, C. F. (2000). *Fingerprinting traffic from static freeway sensors*, University of California Transportation Center, University of California, Berkeley.
- Muñoz, J.C., and Daganzo, C.F (1999). *Experimental observations upstream of an active bottleneck caused by a freeway off-ramp*. PATH working paper, University of California, Berkeley.
- Muñoz, J. C., Daganzo, C. F. (2002). Moving bottlenecks: A theory grounded on experimental observation. *Transportation and Traffic Theory in the 21st Century*. 441-461. Elsevier.
- Nagel, K., Wagner, P., Woesler, R. (2003). Still flowing: Approaches to traffic flow and traffic jam modeling. *Oper. Res.*, 51(5), 681-710.
- Newell, G. F. (1962). Theories of instability in dense highway traffic. *J.Operations Research Society of Japan*, 5(1), 9-54.
- Newell, G. F. (1963). Instability in dense highway traffic, a review. *Proc., Proceedings of the 2nd International Symposium on the Theory of Road Traffic Flow*, 73–83, London.
- Newell, G. (1965). Approximation methods for queues with application to the fixed-cycle traffic light. *SIAM Rev*, 7(2), 223-240.
- Newell, G. (1993). A simplified theory of kinematic waves in highway traffic, part I: General theory. *Transportation Research Part B*. 27(4), 281-287.

- Newell, G. F. (2002). A Simplified CF Theory: A Lower Order Model. *Transportation Research Part B: Methodological*, 36(3), 195-205.
- Noland, R. B., and Quddus, M. A. (2005). Congestion and safety: A spatial analysis of London. *Transportation Research Part A: Policy and Practice*, 39(7-9), 737-754.
- Papageorgiou, M. (1983). *Application of automatic control concepts to traffic flow modeling and control*. Springer, Berlin, Germany.
- Payne, H. J. (1971). Models of freeway traffic and control. *Mathematical Models of Public Systems, No. 28. Simulation Council Proc.*, 51-61, La Jolla, CA.
- Persaud, B., Yagar, S., Brownlee, R. (1998). Exploration of the breakdown phenomenon in freeway traffic. *Transportation Research Record: Journal of the Transportation Research Board*, 1634, 64-69.
- Pipes, L. A. (1953). An operational analysis of traffic dynamics. *J. Appl. Phys.*, 24, 274-281.
- PORTAL. (2009). Portland oregon regional transportation archive listing. <<http://portal.its.pdx.edu/>> (07/01, 2008).
- PTV.(2006). *VISSIM user manual version 4.20*. PTV AG, Karlsruhe, Germany.
- Research Triangle Institute. (1970). *Speed and accidents: Volume I*, US Department of Transportation, Washington, DC.
- Richards, P. I. (1956). Shock waves on the highway. *Oper. Res.*, 4(1), 42-51.
- Schlesselman, J. J., and Stolley, P. D. (1982). *Case-control studies: design, conduct, analysis*. Oxford University Press.
- Schoenhof, M., and Helbing, D. (2007). Empirical features of congested traffic states and their implications for traffic modeling. *Transportation Science*, 41(2), 135-166.
- Shefer, D. (1997). Congestion and safety on highways: Towards an analytical model. *Urban Stud.*, 34(4), 679-692.
- Sheu, J. (1997). *Stochastic estimation of lane-changing probabilities and its applications to incident detection*. Doctoral Dissertation, Department of Civil and Engineering, University of California, Irvine.
- Smith, S. A. (1985). *Freeway data collection for studying vehicle interactions*, Federal Highway Administration; Available through the National Technical Information Service.

- Solomon D. (1964). *Accidents on main rural highways related to speed, driver and vehicle*. US Department of Commerce & Bureau of Public Roads, Washington, DC.
- StataCorp. (2009). Stata: data analysis and statistical software. <<http://www.stata.com/>> (04/01, 2009).
- Sterne, J. A. C., and Smith, G. D. (2001). Sifting the evidence--What's wrong with significance tests? *Phys. Ther.*, 81(8), 1464.
- Synder, D. (1989). Speeding, coordination, and the 55 mph limit: Comment. *Am. Econ. Rev.*, 79(4), 922-925.
- Thiemann, C., Treiber, M., Kesting, A. (2008). Estimating acceleration and lane-changing dynamics based on NGSIM trajectory data. *Transportation Research Record: Journal of the Transportation Research Board*, 2088: 90-101.
- Tilden, C. J., Morris, D. L., Martin, T. M. C., Russell, E. W.(1936). *Motor vehicle speeds on Connecticut highways*, Committee on transportation, Yale University.
- Van Winsum, W., De Waard, D., Brookhuis, K. A. (1999). Lane change maneuvers and safety margins. *Transportation Research Part F: Traffic Psychology and Behaviour*, 2(3), 139-149.
- Wang, C., Coifman, B. (2008). The effect of lane-change maneuvers on a simplified car-following theory. *IEEE Transactions on Intelligent Transportation Systems*, 9(3), 523-535.
- Washington, S., Karlaftis, M., Mannering, F. (2003). *Statistical and econometric methods for transportation data analysis*, Chapman and Hall/CRC Press.
- Westland, D. (1998). Effect of highway geometry on freeway queuing at merge sections. *Proc., Proceedings of the 3rd International Symposium on Highway Capacity*, Edited by R. Rysgaard (Road Directorate, Denmark), 1095–1116.
- Whitham, G.B. (1974). *Linear and nonlinear waves*. Wiley, New York.
- Wiedemann, R. (1991). Modelling of RTI-elements on multi-lane roads. *Advanced Telematics in Road Transport*, 2, 1007-1019.
- Wiedemann, R. (1974). Simulation Des Straßenverkehrsflusses. *Publication Series of the Institute for Traffic (Karlsruhe: University of Karlsruhe)*, 66-68.
- Wilson, R. E. (2008). Mechanisms for spatio-temporal pattern formation in highway traffic models. *Philosophical Transactions-Royal Society of London Series A Mathematical Physical and Engineering Sciences*, 366(1872), 2017.

Li, X., Peng, F., Ouyang Y. Measurement and estimation of traffic oscillation properties. *Transportation Research Part B*, 44(1), 1-14.

Xin, W., Hourdos, J., Michalopoulos, P. G. (2008). Vehicle trajectory collection and processing methodology and its implementation. *the TRB 87th Annual Meeting Compendium of Papers DVD*.

Yeo, H., Skabardonis, A. (2009). Understanding stop-and-go traffic in view of asymmetric traffic theory. *Proceedings of 18th International Symposium of Traffic Theory and Transportation*. Hong Kong. Springer. 99-116.

APPENDIX A

THE FRACTIONAL POLYNOMIAL ANALYSIS

For illustration purposes, the standard deviation of speed is used to explain the fractional polynomial analysis used in this research. For the model, a fractional polynomial of degree m for the standard deviation of speed,

$FP_m\{\text{std. dev. of speed}; (p_1, \dots, p_m)\}$, is defined as

$$\begin{aligned} FP_m\{\text{std. dev. of speed}; (p_1, \dots, p_m)\} \\ = \beta_0 + \beta_1(\text{std. dev. of speed})^{p_1} + \dots + \beta_m(\text{std. dev. of speed})^{p_m} \end{aligned}$$

Where,

$p_1 \leq p_2 \leq \dots \leq p_m$ denote powers (integer or fractional); and

$$(\text{std. dev. of speed})^{(p_j)} = \begin{cases} \ln(\text{std. dev. of speed}) & \text{if } p_j = 0 \\ (\text{std. dev. of speed})^{p_j} & \text{if } p_j \neq 0 \end{cases}$$

For a given degree (e.g., m), the best fitting powers p_1, \dots, p_m are obtained by choosing the model with the smallest deviance. Royston and Altman (1994) provide more technical details of the fractional polynomial analysis.

The simplest functional form of log relative ratio in the standard deviation of speed is linear which corresponds to an FP of first degree ($m = 1$) with a power of one ($p_1 = 1$). The author has tested a second-degree (see Table A1 for the result) and a first-degree (see Table A2) fractional polynomials. Higher-degree fractional polynomials are not considered to keep the model parsimonious.

Table A1 shows that $FP_2\{\text{std. dev. of speed}; (-0.5, 0)\}$ model has the smallest deviance of 253.922 and is better than the model without any variables. However, no significant performance improvement is gained compared

to $FP_1\{std. dev. of speed; (2)\}$ with deviance 258.316 at the 95 percent confidence level (p-value=0.111). Note that p-value in this table is from deviance difference comparing reported model with $m = 2$ model.

Table A1

Summary of the second-degree FP comparisons for std. dev. of speed

Degree of FP, m	Powers	Degree of Freedom	Deviance	Dev. Dif.	P
0 (Not in model)	N/A	0	263.948	10.026	0.04
1	1	1	258.582	4.66	0.198
1	2	2	258.316	4.393	0.111
2	-0.5,0	4	253.922	N/A	N/A

Similarly, according to Table A2, the performance of $FP_1\{std. dev. of speed; (2)\}$ model is not significantly better than the performance of the simple linear model at the 95 percent confidence level (p-value=0.606). Thus, the analysis indicates that it is appropriate to keep the standard deviation of speed linear in the logistic regression model. Note that p-value in this table is from deviance difference comparing reported model with $m = 1$ model.

Table A2

Summary of the first-degree FP comparisons for std. dev. of speed

Degree of FP, m	Powers	Degree of Freedom	Deviance	Dev. Dif.	P
0 (Not in model)	N/A	0	263.948	5.632	0.06
1	1	1	258.582	0.266	0.606
1	2	2	258.316	N/A	N/A

This procedure has been applied to other variables, which resulted in the same conclusion.

APPENDIX B

MODEL EVALUATION RESULTS FOR OTHER CONTROL-TO-CASE RATIOS

Table B1
Model evaluation results (5:1 to 7:1 control-to-case ratios)

control-to-case ratio		5:1		6:1		7:1	
Run	Variables	OR	95% CI	OR	95% CI	OR	95% CI
1	std_spd	1.101	(1.020,1.190)	1.093	(1.018,1.172)	1.084	(1.008,1.165)
2	std_spd	1.062	(0.986,1.143)	1.062	(0.987,1.144)	1.08	(1.002,1.165)
3	std_spd	1.081	(1.003,1.165)	1.116	(1.033,1.206)	1.092	(1.016,1.174)
	avg_occ			1.027	(1.000,1.054)	1.018	(0.997,1.040)
4	std_spd	1.079	(0.997,1.168)	1.075	(0.997,1.158)	1.093	(1.011,1.181)
5	std_spd	1.090	(1.013,1.174)	1.077	(1.002,1.157)	1.088	(1.012,1.170)
6	std_spd	1.069	(0.992,1.150)	1.080	(1.009,1.157)	1.084	(1.012,1.161)
7	std_spd	1.080	(1.005,1.161)	1.077	(1.002,1.157)	1.075	(1.002,1.153)
8	std_spd	1.076	(1.002,1.154)	1.085	(1.012,1.165)	1.066	(0.996,1.140)
9	std_spd	1.071	(0.994,1.154)	1.080	(1.001,1.165)	1.069	(0.994,1.150)
	avg_occ	1.018	(0.996,1.041)	1.018	(0.995,1.042)		
10	std_spd	1.089	(1.009,1.176)	1.088	(1.006,1.177)	1.082	(1.005,1.164)
	avg_occ			1.024	(0.999,1.051)		
11	std_spd	1.121	(1.034,1.215)	1.078	(1.002,1.16)	1.085	(1.011,1.164)
	avg_occ	1.026	(0.999,1.054)				
12	std_spd	1.081	(1.005,1.161)	1.074	(0.999,1.155)	1.085	(1.009,1.167)
	avg_occ			1.017	(0.997,1.038)	1.02	(0.998,1.043)
13	std_spd	1.066	(0.991,1.147)	1.072	(0.991,1.161)	1.068	(0.994,1.147)
	avg_occ					1.020	(0.997,1.044)
	avg_spd			0.981	(0.958,1.004)		
14	std_spd	1.066	(0.990,1.147)	1.074	(0.998,1.156)	1.062	(0.991,1.137)
15	std_spd			1.071	(0.994,1.154)	1.080	(1.001,1.164)
	avg_occ			1.020	(0.995,1.044)	1.028	(1.000,1.056)
16	std_spd	1.096	(1.014,1.185)	1.078	(0.997,1.166)	1.107	(1.019,1.202)
	avg_cnt	0.804	(0.645,1.003)	0.822	(0.662,1.019)		
	avg_spd					0.978	(0.956,1.001)
17	std_spd	1.070	(0.994,1.151)	1.070	(0.996,1.149)	1.087	(1.010,1.170)
	avg_spd	0.980	(0.958,1.003)			0.981	(0.959,1.004)
	avg_occ			1.019	(0.996,1.043)		
18	std_spd	1.083	(1.001,1.173)	1.074	(0.996,1.16)	1.077	(0.997,1.162)
	avg_occ	1.024	(0.997,1.053)	1.021	(0.995,1.05)		
	avg_spd					0.982	(0.960,1.005)
19	std_spd					1.071	(0.998,1.150)
	avg_occ	1.022	(0.997,1.048)	1.020	(0.996,1.044)	1.023	(0.998,1.049)
20	std_spd	1.067	(0.989,1.150)	1.093	(1.013,1.178)	1.075	(1.002,1.153)

Note that the blank cell in this table means that the corresponding variable did not appear in the final model. OR stands for odds ratio.